

The Development of a Typology for Describing Mathematics Assessment Items

Joanne E. Tims

Curtin University of Technology

Abstract

This paper reports the development of a typology for describing mathematics assessment items, with particular reference to aspects of the items which impinge on their gender-fairness. The typology is in two parts - one part to describe the context of the item and the other part to describe the format of the item.

Introduction

This paper describes part of an ongoing research project concerned with gender fair assessment in upper secondary mathematics in Western Australia. It focuses on aspects of the project concerned with the relationship between assessment in mathematics education and the differential performance of males and females. This relationship continues to be a much debated but under-researched area (P. Murphy, 1994).

In trying to understand gender-related patterns of mathematics achievement, a review of the literature reveals that this topic has received a good deal of attention from researchers over the last few decades. Over this time, accepted views about why women underachieve and are under-represented in university science and mathematics faculties have changed considerably. Early research was concentrated in the United States and tended to focus on differences in spatial ability but failed to control for differential course-taking and school-leaving patterns (Willis, 1989). Much of this type of research concluded that females had less mathematical and visual-spatial ability than males. Later research, particularly in Australia, which controlled for differential course taking found no or small gender differences. The questions being asked by researchers in the mathematics and gender field changed from why girls couldn't do mathematics to why they didn't do mathematics. More recently Linn and Hyde (1989), Friedman (1989) and Leder (1992) have provided a new perspective on many of those old studies.

Linn and Hyde (1989) conducted a meta-analysis of research into gender differences in achievement in mathematics and concluded that cognitive gender differences have declined in all areas studied and no longer exist for some content areas, namely spatial visualization and mathematics concepts and computation. In addition Friedman's (1989) meta-analysis examined studies that took place between 1974 to 1987 and concluded that sex differences were very small

and had declined over the years. Leder (1992) reviewed gender difference research published in the *Journal for Research in Mathematics Education*. She found contradictory findings with respect to achievement in mathematics. She suggested that gender differences in mathematics achievement seem to depend on the content and format of the test administered, the age level at which the testing takes place and whether standardized testing or classroom grades are considered. The remainder of this paper will review briefly the research on format and context effects and go on to discuss the development of a framework or typology for describing mathematics assessment items in terms of their gender orientation of format and context.

Previous Research

Item Format Research

Research into differences in performance due to the use of a particular format of assessment task has been reported since the early seventies. In England, R. Murphy (1982) investigated the effect the format of the assessment task has on achievement by analysing the results of boys and girls on 16 separate examinations over four years in a number of different subjects at the end of compulsory schooling. He concluded that males had a significant advantage over females on the objective section of nearly all of the examinations considered. Other studies in England report similar findings (Forrest, 1992). Much research in the United States has reported gender differences in favour of males when multiple choice tests are used. (See for example Mazzeo et al (1991).) In Ireland Bolger and Kellaghan (1990) reported that on the national tests in Mathematics, Irish and English, boys performed significantly better than girls on the multiple choice sections compared to their performance on the free response sections.

R. Murphy (1982) also investigated gender differences in achievement on school based assessments and external examinations and concluded that males tend to perform better on external examinations and females on school based assessments. Ongoing work by Burton (1992) confirms that coursework has a differential effect on performance in mathematics while Rennie and Parker (1991) and Parker (1992a) report similar findings for Western Australian science subjects.

Item Context Research

For the purposes of this paper context will be defined as the non-mathematical setting of a question, and context effects will be defined as the differences in mathematics achievement which occur when different contexts are considered. Research on contextualised mathematics problems has found that personalizing contexts (by including the student's own name and things such as their favourite foods and hobbies) and allowing students to work on preferred contexts can increase performance (Anand & Ross, 1987; L. Murphy & Ross, 1990).

Contextual effects were also found on some earlier work on science assessment. The Assessment of Performance Unit/Department of Education (APU) in the United Kingdom monitored Science process performance in the early 80's. Research on this data (Bransky & Qualter, 1993; Johnson, 1987) showed clearly that science questions with a domestic, social or safety context seemed to enhance girls' performance, whilst any hint of technology in the context depressed girls' performance considerably. Johnson (1987) suggested that one of the reasons for these observed differences is that girls and boys have very different out-of-school experiences, and have therefore, different interests and background knowledge. It has been suggested also (P. Murphy, In press) that contextual clues and features are more significant for girls because of the way in which girls make meaning out of the assessment task. They see the context as important and give consideration to contextual cues which the assessor may not consider relevant. The task as interpreted by the student may thus be quite different from that intended by the assessor. P. Murphy (1994) suggests that this is often not the case for boys, who seem to be able to ignore the context and focus on the assessor's intended task

In summary, it is clear from the research reported that both the format and the context of an assessment task can influence performance of the task-taker in ways which appear to be related to gender. Stronger theoretical frameworks are needed for research in this area however and this paper explores some possibilities in this regard.

Exploring Theoretical Frameworks

It is possible to make sense of gender-related patterns of response to the format and context of assessment tasks by considering the different ways in which boys and girls are socialised in our society. Object relations theory (Chodorow (1989) as cited in Baxter Magolda (1992)) suggests that boys are socialised towards separation and individualisation as they must separate from their mothers in order to develop a masculine identity. This leads to boys struggling with forming relationships. Girls are said to be socialised towards connection, in part because they do not need to differentiate from their mothers in order to establish a feminine identity. This leads to girls struggling with separating from others. The objectivism that has dominated our educational culture for centuries focuses on separating the learner from knowledge and implicitly values a masculine way of separate knowledge above the feminine way of connecting with the learning environment. When applied to assessment, it would seem that girls' socialisation leads them to try to connect with the assessor and the assessment task through the context, whereas boys do not want to connect and are able, therefore, to ignore the context.

In order to examine the gender-fairness of both the context and format of assessment, it is proposed to examine the gender orientation of assessment at both the individual item level and the

whole task level.

The Context of an Assessment Item

At the individual item level, Parker and Rennie (1993) developed a way of classifying the gender orientation of the context of assessment items in physics. This is shown in Table 1 below.

Criteria	Male Orientation	Female Orientation	Allegedly Neutral Orientation	Gender-Inclusive Orientation
Language	Uses he, him, his	Uses she, her, hers	uses they, them, their or role (eg a sprinter....)	uses name of person uses "you"
Portrayal of stereotypes	males in active role, females in passive role	females in active role, males in passive role	genderless people in active role (eg a scientist.....)	both males and females in active and passive roles
Appeal to background experiences	relevant to stereotyped male experiences	relevant to stereotyped female experiences	not relevant to human experiences	relevant to males and females equally
Context	Abstract, decontextualised	human, social	concrete setting	human, social, environmental

These categories were developed from a synthesis of the literature and discussed in workshops with teachers of physics. This classification worked well with physics items and was a useful classification for teachers to use when setting assessment tasks. As yet, it has not been tested to see whether or not items which are male or female-orientated produce biased results. However, considering that the classification was based on previous research which indicated that biased results are produced when females attempt male-orientated items, it seems likely that bias would be found if this classification was tested empirically.

The work of Parker and Rennie with a number of physics teachers was part of a study of teacher implementation of major reforms to the physics curriculum in Western Australian (WA) secondary schools. The new curriculum attempts to humanize physics and make it more relevant to the lives of students. These aims are similar in nature to the aims of the new mathematics curriculum which was examined externally for the first time in 1992 in the Tertiary Entrance Examination (TEE). One of the stated aims of the new WA mathematics curriculum was to provide more relevant mathematics courses, reflecting appropriately up-to-date content and current methodologies (Parker, 1992b).

It was found during the implementation of the the mathematics curriculum that the role of the external exam was emphasized by the mathematics teachers. They felt this would be the most powerful influence on their teaching methods (Parker, 1992b). Hence it was decided to try the Parker and Rennie classification on the first set of TEE mathematics papers. The 1993 annual conference of the Mathematical Association of Western Australia (MAWA) was the ideal place to

access a group of mathematics teachers interested in gender fair assessment. This author conducted a workshop with teachers at the 1993 MAWA conference. At the workshop, the Parker and Rennie classification was used on the 1992 TEE mathematics papers, with a fair degree of success. .

For Discrete Mathematics, the classification worked quite well, as many of the questions had a concrete context. The writers of the examination managed to set nearly every question into a reasonable context, even though some of them were a little contrived.

For Applicable Mathematics it was not so clear, as some of the questions had a concrete context, such as finding the volume of a cylinder, but this context was not connected to human activity. This is the kind of question students may find inaccessible because of its irrelevance to their lives.

The difficulty the teachers had was for the Calculus paper. It seems that for many items in the Calculus TEE, the context was completely devoid of human activity, and the Parker and Rennie classification is based on how the context of the item links the student's experience to the mathematics (or physics). So few questions in the Calculus paper had a concrete setting, let alone a human or social one. The overall tone of the paper was one of inaccessibility. Also, many of the questions were deliberately tricky, making them more inaccessible. Unfortunately the 1993 TEE paper was no better. A public meeting between the examiners for TEE Calculus and teachers of the subject was held in March 1994 at the Secondary Education Authority to discuss the 1993 TEE. Teachers at this meeting expressed the feeling that many of their students had found the 1993 paper tricky and unfriendly. The idea that women prefer learning and working environments where there is connection between human activity and learning is clearly demonstrated in the work of Belenky ,(1986), Shepherd, (1993) and Baxter Magolda (1992). If the context free environment of high school Calculus is considered, there appears to be no connection to the real world. The examiners for TEE Calculus seem to think that Calculus is something which should only happen in the minds of people, not in the real world. Given the teachers' perceptions of the power of the TEE, the implications of this will mean that for at least the foreseeable future, there will be no change in the way Calculus is taught and assessed in WA.

In summary, many mathematics items have a tenuous connection to human activity or are deliberately set in a controversial, confusing or confrontational context. This will often distract the student from the task and can limit the accessibility of an item. The perception of this researcher is that judging the accessibility of the item is difficult for the assessor or even another expert in the subject, but students would have a much better viewpoint from which to judge this attribute. It is anticipated that observations of students completing assessment items and interviews with them after completion of assessment items will help to clarify the sorts of attributes a question has which makes it more or less accessible to students.

Table 2 Gender-Related Patterns of Knowing and Preferred Assessment Styles

Way of Knowing	Absolute		Transitional		Independent		Contextual
Pattern	Received	Mastery	Inter-personal	Impersonal	Inter-individual	Individual	
Gender relation	Female	Male	Female	Male	Female	Male	Both female and male
View of knowledge	Certain - answers always exist		Accept that some knowledge is uncertain		View knowledge as mostly uncertain		Mostly uncertain but some knowledge claims are better than others in certain contexts
Assessment Preference	Offers best opportunity to show knowledge	Helps improve mastery	Takes individual differences into account	fair and practical	Joint process occurring between student and instructor	Based on independent thinking	Flexible - anything as long as it accurately measures competence in a particular context.
Format as used in WA	Assignments with knowledge based objectives	Many tests and assignments of drill and practice type with lots of feedback	Open-ended investigations which provide scope for individuality	Exams and tests with emphasis on understanding and problem solving in real world contexts.	Group projects which allow interaction between instructor and groups	Projects which have a group component and an individual component	Flexible - anything as long as it accurately measures competence in a particular context.

Adapted from Baxter Magolda(1992)

The Format of the Assessment Task

An alternative framework from which to view the gender orientation of assessment is the one offered by Baxter Magolda (1992). Her work describes the development of gender-related patterns of knowing and reasoning in students over four years of college (post-secondary education) in a United States university. Her study followed 135 students from their first year in college for five years, when most had graduated and found employment. She interviewed each student once each year and found that students entering college had a number of gender-related ways of knowing. She indicated that students with different ways of knowing had different preferences for assessment, but she did not explore this in any subject-related way. In the context of previous research, these preferences will be related to methods of assessment used in Western Australian secondary mathematics classes. Interviews with Western Australian secondary students will be conducted to determine whether or not this classification is able to be applied to mathematics assessment in WA.

Table 2 shows the Baxter-Magolda classification of ways of knowing and the proposed operationalised typology of assessment format encompassing the various formats used in mathematics in WA. It is emphasized that at this stage, this framework is developmental. The applications to assessment in mathematics are yet to be tested.

Concluding Comments

The data collection phase of this study is just about to commence, with the major emphasis being on observing and interviewing Year 12 students. The purpose of these observations and interviews is to try to determine what aspects of particular assessment tasks girls and boys perceive differently. At this stage of the research, it is clear that assessment methods used in WA have changed little for the majority of teachers over the last few years, and do not reflect the intended change in teaching methodology which went with the implementation of the new curriculum. What is also clear, is that most teachers are not aware that choice of assessment format or use of context in assessment may have implications for gender-fairness. It is hoped that one of the outcomes of this research will be to raise the awareness of such issues and encourage the use of a wide variety of assessment methods to minimize gender-bias due to method of assessment.

References

- Anand, P. G., & Ross, S. M. (1987). Using computer-assisted instruction to personalize arithmetic materials for elementary school children. Journal of Educational Psychology, 79(1), 72-78.
- Baxter Magolda, M. B. (1992). Knowing and reasoning in college - Gender related patterns in students' intellectual development. San Francisco: Jossey-Bass Publishers.
- Belenky, M. F., Clinchy, B. M., Goldberger, N. R., & Tarule, J. M. (1986). Women's ways of knowing: The development of self, voice and mind. New York: Basic Books.
- Bolger, N., & Kellaghan, T. (1990). Method of measurement and gender differences in scholastic achievement. Journal of Educational Measurement, 27(2), 165-174.
- Bransky, J., & Qualter, A. (1993). Applying physics concepts - uncovering gender differences in assessment of performance unit results. Research in Science and Technological Education, 11(2), 1993.
- Burton, L. (1992). Who assesses who and to what purpose? In M. Stephens & J. Izard (Eds.), Reshaping assessment practices: Assessment in the mathematical sciences under challenge (pp. 1-18). Hawthorn, Vic: Australian Council for Educational Research.
- Forrest, G. M. (1992). Gender differences in school science examinations. Studies in Science Education, 20, 87-122.
- Friedman, L. (1989). Mathematics and the gender gap: A meta-analysis of recent studies on sex differences in mathematical tasks. Review of Educational Research, 59(2), 185-213.
- Johnson, S. (1987). Gender differences in science: Parallels in interest, experience and performance. International Journal of Science Education, 9(467-481).
- Leder, G. C. (1992). Mathematics and gender: Changing perspectives. In D. A. Grouws (Ed.),

- Handbook of Research on Mathematics Teaching and Learning. A project of the National Council of Teachers of Mathematics (pp. 597-622). New York: Macmillan.
- Linn, M. C., & Hyde, J. S. (1989). Gender, mathematics, and science. Educational Researcher, 18(8), 17-27.
- Mazzeo, J., Schmitt, A. P., & Bleistein, C. A. (1991). Do women perform better, relative to men, on constructed-response tests or multiple-choice tests? Evidence from the Advanced Placement Examinations. In Annual meeting of the National Council of Measurement in Education, . Chicago, Illinois:
- Murphy, L. O., & Ross, S. M. (1990). Protagonist gender as a design variable in adapting mathematics story problems to learner interests. Educational Technology, Research and Development, 38(3), 27-37.
- Murphy, P. (1994). Sources of inequity: understanding students' responses to assessment. In Annual Meeting of the American Educational Research Association, . New Orleans, Louisiana:
- Murphy, P. F. (In press). Gender and assessment practice in science. In L. Parker, L. Rennie, & B. Fraser (Eds.), Kluwer
- Murphy, R. J. L. (1982). Sex differences in objective test performance. British Journal of Educational Psychology, 52, 213-219.
- Parker, L. (1992a). Sex differences in performance on external examinations and school-based assessment in science subjects. In Annual Conference of the Western Australian Science Education Association (WASEA), . Perth, Western Australia: Ministry of Education.
- Parker, L. H. (1992b). Curriculum renewal in upper secondary mathematics in Western Australia: A preliminary evaluation. Perth, Western Australia: Key Centre for School Science and Mathematics, Curtin University of Technology.
- Parker, L. H., & Rennie, L. J. (1993). Curriculum reform and gender-fair assessment. In Annual Meeting of the National Association for Research in Science Teaching, Atlanta, GA:
- Rennie, L. J., & Parker, L. H. (1991). Assessment of learning in science: The need to look closely at item characteristics. Australian Science Teachers Journal, 37(4), 56-59.
- Shepherd, L. J. (1993). Lifting the veil. Boston: Shambala Publications Inc.
- Willis, S. (1989). 'Real girls don't do maths' - Gender and the construction of privilege. Geelong: Deakin University Press.